

O I P E

RAW SEQUENCE LISTING

DATE: 09/10/2001

PATENT APPLICATION: US/09/801,852A

TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw

ENTERED

3 <110> APPLICANT: Bristol-Myers Squibb Company
 4 Chiang, Shu-Jen
 5 Jonathan, Basch
 7 <120> TITLE OF INVENTION: DIRECT PRODUCTION OF DESACETYLCEPHALOSPORIN C
 9 <130> FILE REFERENCE: ON0163
 11 <140> CURRENT APPLICATION NUMBER: 09/801,852A
 12 <141> CURRENT FILING DATE: 2001-03-08
 14 <150> PRIOR APPLICATION NUMBER: 60/188,033
 15 <151> PRIOR FILING DATE: 2000-03-09
 17 <160> NUMBER OF SEQ ID NOS: 15
 19 <170> SOFTWARE: PatentIn version 3.0
 21 <210> SEQ ID NO: 1
 22 <211> LENGTH: 1716
 23 <212> TYPE: DNA
 24 <213> ORGANISM: Rhodosporidium toruloides
 26 <400> SEQUENCE: 1
 27 atgctcctta acctcttcac cctcgccctcc ctgctgcca cgctccagct cgcctttgcc 60
 29 tctccgacct ccctcgcccg ccgcacgaac ccaaacgagc cccctcccgt cgtcgacctc 120
 31 ggctacgccc gctaccaagg ctacttgaac gagaccgccc gactctactg gtggcgcgga 180
 33 atccgctacg cctcggetca gcgcttccag gctcctcaga cgcgccgac gcacaaggcc 240
 35 gtccgcaacg cgactgagta tggaccgac tggtggccgg ctacgaggg aaccaacacg 300
 37 accaagggct tgccgcgcc tagcaacagc tcgagcagcg cgcgcagaa acaggcgtcg 360
 39 gaggattgcc tcttctca tgctgttgc cccgcggct cgtgcgagg cgacaatctt 420
 41 cccgtcctcg tctacattca cggagggtgg tacgccttcg gcgatgcgag caccggcagc 480
 43 gactttgccc ccttcacca gcacacggga accaagatgg tcgttgtaaa tctccagtac 540
 45 cgtctcggca gctttggttt cctcgctggc caagccatga aggactacgg tgtaacgaac 600
 47 gccggcttgc ttgaccagca attcgccctt caatgggttc aacagcacgt ctggaagttc 660
 49 ggcggcaacc ccgatcacgt tacgatttgg ggcgagtctg caggcgagc gtccgttatg 720
 51 aaccagatca ttggaacgg cggcaacacc gtcaaggctc tcggtctcaa gaagcccctc 780
 53 ttccacgctg ccacggctc ctccgtcttc ctcccctacc aagccaagta caactcccc 840
 55 ttgcgcgagc tgctctactc ccaactcgctc tcggcgacaa actgcaccaa agccgcctcg 900
 57 tccttcgctt gcctcgaagc tgctcagct gcggcgctcg ctgcggcggg cgtgaagaac 960
 59 tcggcgggcg tcccgttcgg gttttggtcg tatgtcccgg tcgtcgacgg gaccttcttg 1020
 61 actgagcgcg cgtcgctcct tctcgccaag ggcaagaaga acctcaatgg caacctcttc 1080
 63 accgggatca acaacctcga cgaaggattc atattcactg acgccactat tcagaacgac 1140
 65 acgatcagcg accagtcgca gcgcgtctcc cagttcgacc gcctcctcgc cggcctcttc 1200
 67 ccctacatca cctcggagga gcgccaggcc gtcgcgaagc agtaccgat ctccgacgag 1260
 69 ccgtcaaagg gcaacacctt ctctcgcac tcggcgctca tcgcggactc gaccttcgtc 1320
 71 tgcccgaact actggaccgc cgaggcggtc ggctcgtccg cccacaaggg cctcttcgac 1380
 73 tacgcgcgg ctcaccacgc gaccgacaac tcgtactaca tcggctccat ctggaacggc 1440
 75 aagaagtcgg tctcgtccgt ccagtccttc gacggcgcg tcggcggtt catcgagacg 1500
 77 ttcaaccgca acaacaacgc tgccaacaag accatcaacc cttactggcc gacgttcgac 1560
 79 tcgggcaagc agtcctctt caacacgacg acgagggaca ccctctctcc cgccgacccg 1620
 81 cgcacgttg agacttcaag cttgaccgac tttggcacga gccagaagac caagtgcgac 1680
 83 ttctggcggt ggtcaatctc ggtgaacgag ggtctc 1716
 86 <210> SEQ ID NO: 2
 87 <211> LENGTH: 572

RAW SEQUENCE LISTING

DATE: 09/10/2001

PATENT APPLICATION: US/09/801,852A

TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw

88 <212> TYPE: PRT

89 <213> ORGANISM: Rhodosporidium toruloides

91 <400> SEQUENCE: 2

```

93 Met Leu Leu Asn Leu Phe Thr Leu Ala Ser Leu Ala Ala Thr Leu Gln
94 1      5      10      15
96 Leu Ala Phe Ala Ser Pro Thr Ser Leu Val Arg Arg Thr Asn Pro Asn
97      20      25      30
99 Glu Pro Pro Pro Val Val Asp Leu Gly Tyr Ala Arg Tyr Gln Gly Tyr
100     35     40     45
102 Leu Asn Glu Thr Ala Gly Leu Tyr Trp Trp Arg Gly Ile Arg Tyr Ala
103     50     55     60
105 Ser Ala Gln Arg Phe Gln Ala Pro Gln Thr Pro Ala Thr His Lys Ala
106 65     70     75     80
108 Val Arg Asn Ala Thr Glu Tyr Gly Pro Ile Cys Trp Pro Ala Ser Glu
109      85     90     95
111 Gly Thr Asn Thr Thr Lys Gly Leu Pro Pro Pro Ser Asn Ser Ser Ser
112      100    105    110
114 Ser Ala Pro Gln Lys Gln Ala Ser Glu Asp Cys Leu Phe Leu Asn Val
115     115    120    125
117 Val Ala Pro Ala Gly Ser Cys Glu Gly Asp Asn Leu Pro Val Leu Val
118     130    135    140
120 Tyr Ile His Gly Gly Gly Tyr Ala Phe Gly Asp Ala Ser Thr Gly Ser
121 145     150    155    160
123 Asp Phe Ala Ala Phe Thr Lys His Thr Gly Thr Lys Met Val Val Val
124      165    170    175
126 Asn Leu Gln Tyr Arg Leu Gly Ser Phe Gly Phe Leu Ala Gly Gln Ala
127     180    185    190
129 Met Lys Asp Tyr Gly Val Thr Asn Ala Gly Leu Leu Asp Gln Gln Phe
130     195    200    205
132 Ala Leu Gln Trp Val Gln Gln His Val Ser Lys Phe Gly Gly Asn Pro
133     210    215    220
135 Asp His Val Thr Ile Trp Gly Glu Ser Ala Gly Ala Gly Ser Val Met
136 225     230    235    240
138 Asn Gln Ile Ile Ala Asn Gly Gly Asn Thr Val Lys Ala Leu Gly Leu
139      245    250    255
141 Lys Lys Pro Leu Phe His Ala Ala Ile Gly Ser Ser Val Phe Leu Pro
142     260    265    270
144 Tyr Gln Ala Lys Tyr Asn Ser Pro Phe Ala Glu Leu Leu Tyr Ser Gln
145     275    280    285
147 Leu Val Ser Ala Thr Asn Cys Thr Lys Ala Ala Ser Ser Phe Ala Cys
148     290    295    300
150 Leu Glu Ala Val Asp Ala Ala Ala Leu Ala Ala Ala Gly Val Lys Asn
151 305     310    315    320
153 Ser Ala Ala Phe Pro Phe Gly Phe Trp Ser Tyr Val Pro Val Val Asp
154     325    330    335
156 Gly Thr Phe Leu Thr Glu Arg Ala Ser Leu Leu Leu Ala Lys Gly Lys
157     340    345    350
159 Lys Asn Leu Asn Gly Asn Leu Phe Thr Gly Ile Asn Asn Leu Asp Glu
160     355    360    365

```

RAW SEQUENCE LISTING

DATE: 09/10/2001

PATENT APPLICATION: US/09/801,852A

TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw

```

162 Gly Phe Ile Phe Thr Asp Ala Thr Ile Gln Asn Asp Thr Ile Ser Asp
163      370      375      380
165 Gln Ser Gln Arg Val Ser Gln Phe Asp Arg Leu Leu Ala Gly Leu Phe
166 385      390      395      400
168 Pro Tyr Ile Thr Ser Glu Glu Arg Gln Ala Val Ala Lys Gln Tyr Pro
169      405      410      415
171 Ile Ser Asp Ala Pro Ser Lys Gly Asn Thr Phe Ser Arg Ile Ser Ala
172      420      425      430
174 Val Ile Ala Asp Ser Thr Phe Val Cys Pro Thr Tyr Trp Thr Ala Glu
175      435      440      445
177 Ala Phe Gly Ser Ser Ala His Lys Gly Leu Phe Asp Tyr Ala Pro Ala
178      450      455      460
180 His His Ala Thr Asp Asn Ser Tyr Tyr Ile Gly Ser Ile Trp Asn Gly
181 465      470      475      480
183 Lys Lys Ser Val Ser Ser Val Gln Ser Phe Asp Gly Ala Leu Gly Gly
184      485      490      495
186 Phe Ile Glu Thr Phe Asn Pro Asn Asn Asn Ala Ala Asn Lys Thr Ile
187      500      505      510
189 Asn Pro Tyr Trp Pro Thr Phe Asp Ser Gly Lys Gln Leu Leu Phe Asn
190      515      520      525
192 Thr Thr Thr Arg Asp Thr Leu Ser Pro Ala Asp Pro Arg Ile Val Glu
193      530      535      540
195 Thr Ser Ser Leu Thr Asp Phe Gly Thr Ser Gln Lys Thr Lys Cys Asp
196 545      550      555      560
198 Phe Trp Arg Gly Ser Ile Ser Val Asn Ala Gly Leu
199      565      570

```

201 <210> SEQ ID NO: 3

202 <211> LENGTH: 2220

203 <212> TYPE: DNA

204 <213> ORGANISM: Rhodosporidium toruloides

206 <400> SEQUENCE: 3

```

207 ggatccaccc gaactctgtc ccgctttctg gctttcttcc ttgctgtcgc cccatgcgct      60
209 ttcccgactc gccgccatgc tccttaacct cttcaccctc gcctccctcg ctgcgacgct      120
211 ccagctcgcc tttgcctctc cgacctccct cgtccgcgcg acgaacccaa acgagccccc      180
213 tcccgctcgtc gacctcgggt acgcccgcga ccaaggctac ttgaacgaga ccgccggact      240
215 ctactggtgg cgcggaatcc gctacgcctc ggctcagcgc ttccaggctc ctgagacgcc      300
217 cgcgacgcac aaggccgtcc gcaacgcgac tgagtatgga ccgatctgtt ggccggctag      360
219 cgagggaacc aacacgacca agggcttgcc gccgcctagc aacagctcga gcagcgcgcc      420
221 gcagaaacag gcgtcggagg attgcctctt cctcaatgtc gttgcccccg ccggctcgtg      480
223 cgaggggcgac aatcttcccc tctcgtctca cattcacgga ggtggctacg ccttcggcga      540
225 tgcgagcacc ggcagcgact ttgccgcctt caccaagcac acgggaacca agatggtcgt      600
227 tgtaaattctc cagtaccgtc tcggcagctt tggtttcctc gctggccaag ccatgaagga      660
229 ctacggtgta acgaacgccg gcttgcttga ccagggtgagt ttcccgcgat ataccgccc      720
231 acctttcgac toatgctgac gcctctcccc ctgcgagcaa ttcgcccttc aatgggttca      780
233 acagcacgctc tcgaagtctc gcggaacccc cgatcacgtt acgatttggg gcgagtctgc      840
235 aggcgcaggg tccgttatga accagatcat tgcgaacgtg agccaccga accgatctcc      900
237 agccgacttt cccccccccc cccccccgcg tgacctccct cgtcttgtag ggcggaaca      960
239 ccgtcaaggc tctcgggtctc aagaagcccc tcttccacgc tgccatcggc tcctccgtct      1020
241 tctcccccta ccaagccaag tacaactccc ccttcgccga gctgctctac tcccaactcg      1080

```

RAW SEQUENCE LISTING

DATE: 09/10/2001

PATENT APPLICATION: US/09/801,852A

TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw

```

243 tctcggcgac aaactgcacc aaagccgcct cgtccttcgc ttgcctcgaa gctgtcgacg 1140
245 ctgcggcgct cgtcgcggcg ggcgtgaaga actcggcggc gttcccgttc gggttttggt 1200
247 cgtatgtccc ggtcgtcgac gggaccttct tgactgagcg cgcgtcgctc cttctcgcca 1260
249 agggcaagaa gaacctcaat ggcgtgcgtg gcgagctttc gagtgcctca ggatctcgct 1320
251 gacactgtcg accggctcgc agaacctctt caccgggacg aacaacctcg acgaagatga 1380
253 gttcccgtcg acggctctgt tcgccagcg agactgactt gttcttttgc gaagattacg 1440
255 attcatattc actgacgcca ctattcagaa cgacacgacg agcgaccagt cgcagcgcg 1500
257 ctcccagttc gaccgcctcc tcgccggcct cttcccctac atcacctcgg aggagcgcca 1560
259 ggccgtcgcg aagcagtaac cgatctccga cgcgcgcgtc aagggcaaca cttctctcg 1620
261 catctcggcc gtcacgcggg actcgacctt cgtgtgcgtt ccccgtcgtc ttctccgagt 1680
263 attccgctga cttcccgtt gccgcagct gcccgacctc ctggaccgcc gaggcgttcg 1740
265 gctcgtccgc ccacaagggc ctcttcgact acgcgcggcg tcaccacgcg accgacaact 1800
267 cgtactacat cggctccatc tggaacggca agaagtcggt ctcgtccgtc cagtcttcg 1860
269 acggcgcgct cggcggttc atcgagacgt tcaaccgaa caacaacgct gccaacaaga 1920
271 ccatacaacc ttactggcgg acgtctcact cgggcaagca gctcctcttc aacacgacga 1980
273 cgagggacac cctctctccc gccgaccgcg gcacgttgga gacttcaagc ttgaccgact 2040
275 ttggcacgag ccagaagacc aagtgcgact tctggcgtgg gtcaatctcg gtgaacgcgg 2100
277 gtctctaggg gtctttcctt ccgacttctt tcgttctttc gttgtttatt cttgcagttc 2160
279 cgttgtatcg gccattcgtg cgtgtagctc actcgagtat agacgttggc aagtgcgaaa 2220

```

282 <210> SEQ ID NO: 4

283 <211> LENGTH: 544

284 <212> TYPE: PRT

285 <213> ORGANISM: Rhodosporidium toruloides

287 <400> SEQUENCE: 4

```

289 Thr Asn Pro Asn Glu Pro Pro Pro Val Val Asp Leu Gly Tyr Ala Arg
290 1 5 10 15
292 Tyr Gln Gly Tyr Leu Asn Glu Thr Ala Gly Leu Tyr Trp Trp Arg Gly
293 20 25 30
295 Ile Arg Tyr Ala Ser Ala Gln Arg Phe Gln Ala Pro Gln Thr Pro Ala
296 35 40 45
298 Thr His Lys Ala Val Arg Asn Ala Thr Glu Tyr Gly Pro Ile Cys Trp
299 50 55 60
301 Pro Ala Ser Glu Gly Thr Asn Thr Thr Lys Gly Leu Pro Pro Pro Ser
302 65 70 75 80
304 Asn Ser Ser Ser Ser Ala Pro Gln Lys Gln Ala Ser Glu Asp Cys Leu
305 85 90 95
307 Phe Leu Asn Val Val Ala Pro Ala Gly Ser Cys Glu Gly Asp Asn Leu
308 100 105 110
310 Pro Val Leu Val Tyr Ile His Gly Gly Gly Tyr Ala Phe Gly Asp Ala
311 115 120 125
313 Ser Thr Gly Ser Asp Phe Ala Ala Phe Thr Lys His Thr Gly Thr Lys
314 130 135 140
316 Met Val Val Val Asn Leu Gln Tyr Arg Leu Gly Ser Phe Gly Phe Leu
317 145 150 155 160
319 Ala Gly Gln Ala Met Lys Asp Tyr Gly Val Thr Asn Ala Gly Leu Leu
320 165 170 175
322 Asp Gln Gln Phe Ala Leu Gln Trp Val Gln Gln His Val Ser Lys Phe
323 180 185 190
325 Gly Gly Asn Pro Asp His Val Thr Ile Trp Gly Glu Ser Ala Gly Ala

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001

TIME: 11:23:03

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw

```

326          195          200          205
328 Gly Ser Val Met Asn Gln Ile Ile Ala Asn Gly Gly Asn Thr Val Lys
329          210          215          220
331 Ala Leu Gly Leu Lys Lys Pro Leu Phe His Ala Ala Ile Gly Ser Ser
332 225          230          235          240
334 Val Phe Leu Pro Tyr Gln Ala Lys Tyr Asn Ser Pro Phe Ala Glu Leu
335          245          250          255
337 Leu Tyr Ser Gln Leu Val Ser Ala Thr Asn Cys Thr Lys Ala Ala Ser
338          260          265          270
340 Ser Phe Ala Cys Leu Glu Ala Val Asp Ala Ala Ala Leu Ala Ala Ala
341          275          280          285
343 Gly Val Lys Asn Ser Ala Ala Phe Pro Phe Gly Phe Trp Ser Tyr Val
344          290          295          300
346 Pro Val Val Asp Gly Thr Phe Leu Thr Glu Arg Ala Ser Leu Leu Leu
347 305          310          315          320
349 Ala Lys Gly Lys Lys Asn Leu Asn Gly Asn Leu Phe Thr Gly Ile Asn
350          325          330          335
352 Asn Leu Asp Glu Gly Phe Ile Phe Thr Asp Ala Thr Ile Gln Asn Asp
353          340          345          350
355 Thr Ile Ser Asp Gln Ser Gln Arg Val Ser Gln Phe Asp Arg Leu Leu
356          355          360          365
358 Ala Gly Leu Phe Pro Tyr Ile Thr Ser Glu Glu Arg Gln Ala Val Ala
359          370          375          380
361 Lys Gln Tyr Pro Ile Ser Asp Ala Pro Ser Lys Gly Asn Thr Phe Ser
362 385          390          395          400
364 Arg Ile Ser Ala Val Ile Ala Asp Ser Thr Phe Val Cys Pro Thr Tyr
365          405          410          415
367 Trp Thr Ala Glu Ala Phe Gly Ser Ser Ala His Lys Gly Leu Phe Asp
368          420          425          430
370 Tyr Ala Pro Ala His His Ala Thr Asp Asn Ser Tyr Tyr Ile Gly Ser
371          435          440          445
373 Ile Trp Asn Gly Lys Lys Ser Val Ser Ser Val Gln Ser Phe Asp Gly
374          450          455          460
376 Ala Leu Gly Gly Phe Ile Glu Thr Phe Asn Pro Asn Asn Asn Ala Ala
377 465          470          475          480
379 Asn Lys Thr Ile Asn Pro Tyr Trp Pro Thr Phe Asp Ser Gly Lys Gln
380          485          490          495
382 Leu Leu Phe Asn Thr Thr Thr Arg Asp Thr Leu Ser Pro Ala Asp Pro
383          500          505          510
385 Arg Ile Val Glu Thr Ser Ser Leu Thr Asp Phe Gly Thr Ser Gln Lys
386          515          520          525
388 Thr Lys Cys Asp Phe Trp His Gly Ser Ile Ser Val Asn Ala Gly Leu
389          530          535          540
391 <210> SEQ ID NO: 5
392 <211> LENGTH: 15
393 <212> TYPE: PRT
394 <213> ORGANISM: Rhodosporidium toruloides
396 <400> SEQUENCE: 5
398 Thr Asn Pro Asn Glu Pro Pro Pro Val Val Asp Leu Gly Tyr Ala

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/801,852A

DATE: 09/10/2001

TIME: 11:23:04

Input Set : A:\ON0163.ST25Rev.txt

Output Set: N:\CRF3\09102001\I801852A.raw